

# Subhadip Mitra

*Data & AI Executive*

Singapore | [contact@subhadipmitra.com](mailto:contact@subhadipmitra.com)

[LinkedIn](#) | [GitHub](#) | [Website](#)

*Last updated: April 1, 2026*

## Professional Summary

---

I build the practices, teams, and systems that turn AI from research into enterprise reality. 15+ years driving enterprise transformation across data engineering, AI systems, and organizational leadership. Head of Data & Analytics for Google Cloud Southeast Asia; built the practice from zero to ~\$X0M in delivery value across 7 countries, overseeing a ~\$X0M regional portfolio as Site Lead. Member of Google's *delta* innovation team with 8 technical disclosures and peer-reviewed research published at ICLR, spanning distributed systems, LLM inference, and agentic AI. Previously Principal Engineer at Standard Chartered Bank and Solution Architect at Teradata, with early career at Microsoft on Windows Kernel.

## Professional Experience

---

**Google Cloud - Professional Services Organization: Head of Data & Analytics, Southeast Asia**  
— Site Lead, PSO Southeast Asia *November 2021 – Present*

Progressed from Senior Technical Solutions Consultant to Head of Data & Analytics (Year 2) to additionally Site Lead for all PSO practices in Southeast Asia (Year 3). Member of *delta*, Google Cloud Consulting's innovation team of AI and engineering experts delivering enterprise transformation from vision to production.

### Leadership & Strategy:

- Built D&A practice from zero to ~\$X0M cumulative delivery value across 6 SEA countries, with delivery extending to Australia, Japan, India, and Korea. As Site Lead, oversaw ~\$X0M regional portfolio across all 7 PSO practices, achieving 97% CSAT and 100% annual revenue targets.
- Created Practice's Data Strategy competency from scratch, generating \$X0M+ in pursuit pipeline within 2 years through 14 executive engagements. Stood up the first AI COE for a major Thai bank and first Data Monetization platform for a leading Indonesian telco.
- Rescued a strategic SEA account facing platform exodus by designing a multi-agent AI solution accelerator, converting a potential departure into a \$X0M+ 3-year commitment. Led recovery of a \$X0M at-risk cross-regional engagement called upon by senior leadership.

### Enterprise Delivery & Innovation:

- Delivered first-of-its-kind ML notebook migration (30K+) for a major bank using a custom multi-agent system, unblocking GCP adoption. Led GenAI reconciliation for a major airline (replicated across JAPAC), petabyte-scale modernizations for Fortune 500 clients, and a sovereign GenAI program across 20+ government agencies.
- Designed and launched JAPAC's Agent Program Office (APO), operationalizing AI agents across 5 practices and 6 sub-regions, sponsored by MD of PSO JAPAC. Made agentic assessment mandatory for all Deal Reviews, established Agent Champions network, launched 7 pilot projects. Targeting 10% efficiency across 120+ person org, positioning JAPAC as global benchmark for agent-driven delivery.

### People & Cross-functional Leadership:

- Built hiring pipelines across Indonesia, India, Australia, and Greater China. Mentored practitioners through Google's internal mentorship program and the Skills Ignition national traineeship (IMDA Singapore partnership).
- Partnered with Sales leadership across SEA on go-to-market strategy and account planning. Coordinated across all JAPAC practice leads to define agent-building ownership (targeting 70-80% practice-built agents) and integrate agentic efficiency into existing delivery governance.

**Standard Chartered Bank: Principal Engineer - Data & Analytics Transformation** *July 2019 – November 2021*

Led design and development of retail bank's data & analytics platform serving 11 markets, 100+ systems, and 1200+ users.

- Developed self-service ML Workbench with 500+ production models, reducing deployment time from 6 months to 1 week
- Architected MarTech strategy driving 30% increase in customer acquisition through data-driven personalization
- Created credit risk models over 15,000+ named entities leveraging news trends and social signals, reducing potential losses by \$XM

**Think Big Analytics (a Teradata company): Principal Data Engineer / Solution Architect** *May 2017 – July 2019*

Architected enterprise-scale data solutions for Fortune 500 clients across APAC.

- Designed 5 global data lakes with ETL pipelines handling 1.2 PB/hour and 40K daily files
- Engineered real-time platform processing 2.5M events/second, improving Ad campaign responsiveness by 80%
- Built ML fraud detection system achieving 60% fewer false positives and 25% higher detection rates, resulting in \$XM savings

**UTU Singapore: Technical Lead**

*April 2016 – May 2017*

Led technical strategy and Thailand market expansion for fintech payments platform.

- Built high-performance payment/rewards systems: 99.99% uptime, sub-100ms latency
- Defined technology roadmaps and bank integration strategy, working directly with C-level executives and banking partners

**Truckaurbus: Founder & CTO**

*March 2014 – March 2016*

Founded B2B commercial vehicle marketplace connecting fleet operators with OEMs and financing partners.

- Led end-to-end technical strategy and product development; scaled to 15 cities with 25+ OEM and bank partnerships

**Microsoft: Senior SDE**

*January 2010 – February 2014*

Core OS development in the Windows & Devices Group.

- Windows Kernel development for Windows 7, Windows 8, and Windows Server 2012 R2
- Azure Machine Learning implementations and CDN architecture optimization

## Research & Open Source Engineering

---

### Spark LLM Eval - Distributed Evaluation Framework

Enterprise-scale LLM evaluation on Apache Spark addressing the gap in statistically rigorous model assessment. Bootstrap confidence intervals, paired significance tests, multi-provider support (OpenAI, Anthropic, Gemini), Delta Lake integration.

[GitHub](#) · [Blog](#)

### LLM Inference Efficiency Research

Tackling the memory-bandwidth bottleneck in LLM inference through speculative decoding (6 techniques, 2-3x speedup) and custom Triton kernels (RMSNorm 8.1x faster, 88% peak A100 bandwidth). Supports CUDA, Apple Silicon, and CPU.

[GitHub](#) · [GitHub](#)

*Google Technical Disclosure - Pending*

### AI Metacognition Toolkit

Activation-level detection of sandbagging and deception in LLMs. Linear probes achieve 90-96% accuracy across Mistral, Gemma, and Qwen; steering vectors reduce sandbagging by 20% without retraining. Published

on PyPI.

[PyPI](#) · [GitHub](#) · [Docs](#) · [Blog](#)

### **Steering Vectors for Agent Behavior Control**

Runtime control of LLM agent behaviors through activation steering - 65% uncertainty detection on ambiguous queries while maintaining 100% confidence on factual ones. LangChain integration, tested on Mistral-7B, Gemma-2-9B, Qwen3-8B.

[GitHub](#) · [Blog](#)

### **CatchMe - Intelligent Trust Engine**

Industry-agnostic agentic AI for enterprise trust decisions. Automated Pattern Learning System (APLS) with cascade routing achieving 86% cost reduction and sub-50ms latency.

[Website](#)

## **Publications & Technical Disclosures**

---

### **Quality-Diversity Evolution for Discovering Diverse Vulnerabilities in LLM Safety**

*ICLR 2026 Workshop AIWILD, May 2026*

### **Spark-LLM-Eval: A Distributed Framework for Statistically Rigorous Large Language Model Evaluation**

*arXiv, January 2026*

### **Field-Theoretic Memory for AI Agents: Continuous Dynamics for Context Preservation**

*arXiv, January 2026*

### **Automated Rule Generation for Tiered Systems Using Multi-Stage Failure Learning**

*Google, Technical Disclosure Commons, January 2026*

### **Predictive Cost-Benefit Routing for Multi-Tier Risk Decisioning Systems**

*Google, Technical Disclosure Commons, January 2026*

### **UPIR: Automated Synthesis and Verification of Distributed Systems**

*Google, Technical Disclosure Commons, November 2025*

### **ETLC: A Context-First Approach to Data Processing in the Generative AI Era**

*Google Cloud, May 2025*

### **Field-Theoretic Context System (FTCS)**

*Google, Technical Disclosure Commons, May 2025*

### **ARTEMIS - Adaptive Multi-agent Debate Framework**

*Google, Technical Disclosure Commons, January 2025*

### **Data Monetization Strategy for Enterprises**

*BITS Pilani, December 2023*

### **OConsent: Open Consent Protocol for Privacy and Consent Management with Blockchain**

*BITS Pilani, December 2021*

## **Education**

---

**MBA, Business Analytics**

Birla Institute of Technology and Science, Pilani

**MTech, Software Systems**

Birla Institute of Technology and Science, Pilani

## **Technical Skills**

---

**Data Engineering & AI:** Apache Spark, Apache Kafka, Delta Lake, Apache Iceberg, PyTorch, LangChain, LangGraph, Google ADK, Multi-Agent Systems, RAG Architecture, MLflow

**Cloud & Infrastructure:** Google Cloud Platform, Vertex AI, GKE, Terraform, Kubernetes

**Programming & Systems:** Python, SQL, Scala, Triton, CUDA